Benefits of Imperfect Conflict Resolution Advisory Aids for Future Air Traffic Control

Trapsilawati, Fitri; Wickens, Christopher D.; Qu, Xingda; Chen, Chun-Hsien

2016

Trapsilawati, F., Wickens, C. D., Qu, X., & Chen, C.-H. (2016). Benefits of imperfect conflict resolution advisory aids for future air traffic control. *Human factors*, 58(7), 1007-1019.

<https://hdl.handle.net/10356/85652>

<https://doi.org/10.1177/0018720816655941>

© 2016 Human Factors and Ergonomics Society (HFES). This is the author created version of a work that has been peer reviewed and accepted for publication by Human Factors, Human Factors and Ergonomics Society (HFES). It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1177/0018720816655941>].

Downloaded on 26 Apr 2023 17:22:45 SGT

Benefits of imperfect conflict resolution advisory aids for future air traffic control

(Accepted-in Production)

Fitri Trapsilawati¹, Christopher D. Wickens², Xingda Qu³, and Chun-Hsien Chen¹

¹ School of Mechanical and Aerospace Engineering,
Nanyang Technological University, Singapore

² Alion Science and Technology, Boulder, Colorado, United States

³ Institute of Human Factors and Ergonomics,
College of Mechatronics and Control Engineering
Shenzhen University, Shenzhen, China

Running head: Conflict Resolution Automation for ATC

Topic choice: Aviation and Aerospace

Manuscript type: Research Article

Word count of text: 4519

Acknowledgements:

Corresponding author: Xingda Qu, PhD; Address: 3688 Nanhai Avenue, Shenzhen, Guangdong
Province 518060, China; Phone: +86-755-86965716; Email: quxd@szu.edu.cn.

Funding sources: Research was supported by Civil Aviation Authority of Singapore (CAAS) and Air Traffic Management Research Institute (ATMRI), Nanyang Technological University (NTU), Singapore, project reference ATMRI:2014-R5-CHEN.

Abstract

Objective: To examine the human-automation interaction issues and the interacting factors in the context of conflict detection and resolution advisory (CRA) system.

Background: The issues of imperfect automation in air traffic control have been well documented in previous studies, particularly in conflict alerting systems. The extent to which the prior findings can be applied to an integrated conflict detection and resolution system in future air traffic control (ATC) remains unknown.

Methods: Twenty-four participants were evenly divided into two groups corresponding to a medium and a high traffic density conditions, respectively. In each traffic density condition, participants were instructed to perform simulated ATC tasks under four automation conditions including reliable, unreliable with short Time Allowance to Secondary conflict (TAS), unreliable with long TAS, and manual conditions. Dependent variables accounted for conflict resolution performance, workload, situation awareness, trust in and dependence on the CRA aid, respectively.

Results: Imposing the CRA automation did increase performance and reduce workload as compared to manual performance. The CRA aid did not decrease situation awareness. The benefits of the CRA aid were manifest even when it was imperfectly reliable and were apparent across traffic loads. In the unreliable blocks, trust in the CRA was degraded but dependence was not influenced, yet the performance was not adversely affected.

Conclusion: The use of CRA aid would benefit ATC operations across traffic densities.

Application: CRA aid offers benefits across traffic densities, regardless of its imperfection, as long as its reliability level is set above the threshold of assistance, suggesting its application for future ATC.

Keywords: automation; human-automation interaction; air traffic control; conflict resolution; mental workload; situation awareness

Précis

CRA aid is a promising concept of automation aid for the future ATC system. The current study provides empirical evidence of the benefits of CRA aid in high air traffic environment, notwithstanding its potential imperfection. These findings highlight the need for the provision of CRA aid for future ATC operations.

INTRODUCTION

Over the last two decades, many researchers have focused on the implications of imperfections in automation (Li, et. al, 2014; Manzey, Gérard, & Wiczorek, 2014; Mosier & Skitka, 1996; Onnasch, Wickens, et al., 2014; Parasuraman & Manzey, 2010; Parasuraman, Molloy, & Singh, 1993; Parasuraman, Sheridan, & Wickens, 2000; Rovira, McGarry, & Parasuraman, 2007; Rovira & Parasuraman, 2010; Wickens, et al., 2015; Yeh, et al., 2003). Research on human-automation issues in air traffic conflict alerting systems remain quite relevant to the issues of imperfection in automation (Metzger & Parasuraman, 2005; Wickens, Rice, et al., 2009). Much of this research has focused on the existing conflict alert (CA) system in ATC facilities, a system that makes inferences about a pending loss of separation between aircraft, and warns Air Traffic Controllers (ATCOs) accordingly. The existence of imperfections in such warning systems via their high false alarm rate has been well documented (Wickens, Rice, et al., 2009).

To date however, active controllers have not been served by the automated **conflict resolution advisory** (CRA) systems; a system that not only detects conflicts but also recommends ATCOs maneuvers to eliminate the conflict (Prevot, et al., 2012; Trapsilawati, et al., 2015) although such systems do exist on the flight deck. Given the fact of increasing air traffic (Airbus, 2013), one of the main challenges in ATC is the higher probability of air traffic conflicts. In fact, in research the CRA has been found to be able to improve ATCOs' situation awareness (SA) as well as help them develop more accurate and faster conflict resolutions (Prevot et al., 2012; Trapsilawati et al., 2015).

Returning to the simpler ATC conflict alert system, while the automation imperfections here are well documented (Wickens, Rice, et al., 2009), it is apparent that even such imperfect

81 automation **can** assist ATCOs in conflict detection and understanding relative to totally unaided
82 performance (Metzger & Parasuraman, 2005). Indeed a meta-analysis of studies of these, and
83 other automation-supported detection tasks reveals that the reliability of such automation can be
84 as low as 70-75% and still assist human-system performance relative to human-only performance
85 (Wickens & Dixon, 2007). This is particularly true when the workload imposed on human's
86 cognition is high.

87 High workload influences the tendency to depend on, to agree with and to accept more
88 automation aid (Wickens & Dixon, 2007), particularly in a high traffic environment (Prevot, et
89 al., 2012; Westin, Borst, & Hilburn, 2013) where ATCOs have limited time to perform ATC
90 simultaneous tasks (Vossen, Hoffman, & Mukherjee, 2012). However, high task load does not
91 necessarily affect trust. Trust is affected more critically by automation reliability (Wickens, et al.
92 2013) and is generally rated lower with lower automation reliability (Metzger & Parasuraman,
93 2005).

94 In the current research, we use the CRA aid that integrates the conflict detection and
95 resolution system as a tool for investigating the above issues in human-automation interaction
96 (HAI). This choice allows us to examine the imperfections in the CRA system, in a manner that
97 has rarely been done within the very small body of research that has examined controller
98 performance with the CRA (Cabrall et al., 2014; Prevot, Homola, & Mercer, 2008; Prevot et al.,
99 2012; Trapsilawati et al., 2015). Only Trapsilawati et al. (2015) examined this issue of imperfect
100 CRA and those authors did indeed find that the imperfect CRA aid assisted ATCOs relative to
101 manual performance. However they did not vary task load to examine if this factor amplified
102 CRA aid benefits and imperfection costs.

In the current experiment we had a much higher level (2X and 3X) of traffic load, where dependence on automation might be expected to be considerably higher, thereby amplifying both its benefits and its costs when it errs (Wickens & Dixon, 2007). Thus the current study extends the general paradigm employed by Trapsilawati et al. to examine these issues. The issue of imperfect automation is thoroughly addressed in recent work by Onnash et al (2014) and by Wickens, Clegg, et al. (2015). Here the distinction is made between the overall level of performance supported by automation that is above around 75% correct on the one hand (aggregating both automation correct and automation failure trials), and the specific performance on the infrequent occasions (e.g., 25%) when automation fails on the other. The former value indicates overall automation assistance relative to unaided manual performance (the weighted average of correct and automation-error trials). This is the value employed by Wickens & Dixon to define the threshold of automation assistance. However the performance on the rare automation failure trial suffers. This degradation on the automation failure trial results, in part, because of the operator increases dependence on automation during the frequent trials of its correct operation, becomes partially out of the loop (a complacency or “automation bias” effect), and hence loses situation awareness, as revealed by the meta-analysis of Onnasch et al. (2014). As a consequence there is a less fluid intervention when automation **does** fail; and this is particularly prominent the first time automation fails (Wickens, Clegg, et al., 2015).

To examine these issues, we had participants with ATC experience control traffic manually and also use a CRA tool that was either fully or partially reliable under low and high levels of workload (traffic density). When the CRA was unreliable, we imposed either a short or long Time Allowance to Secondary conflict (TAS). We examined a variety of performance and cognitive variables both when the CRA aid worked perfectly and imperfectly.

Based on prior research described above, seven hypotheses were offered (Table 1).

Table 1. Research hypotheses

Hyp. No	Hypotheses
H1	Imposing the CRA automation will increase both performance and reduce workload (when compared to fully manual conflict resolution), given that controllers use (depend upon) the automation.
H2	The benefits in (1) will be amplified under high traffic load, when cognitive resources were scarcer.
H3	The benefits of the CRA automation will be manifest even when it was imperfectly reliable, given that its reliability was greater than the threshold of automation assistance between 0.70 & 0.75 (Wickens & Dixon, 2007).
H4	CRA automation, whether perfect or imperfect will decrease situation awareness (SA) of the traffic situation.
H5	The effects in (4) will be amplified under high workload (high traffic load).
H6	After automation has failed, the trust in the CRA will decline but the dependence will not decline, due to the high traffic load.
H7	Shorter TAS should decrease situation awareness, increase workload and increase dependence on automation.

METHODS

Participants

Twenty four participants aged 19 to 34 years (M= 24 years, SD= 3.33 years) were recruited. The participants consisted of 22 students and 2 professional ATCOs in Singapore. The students majored in Aerospace and Aeronautical Engineering from local institutions and had at least two weeks ATC simulator experience during their course. An interview was conducted before the study to make sure that all these participants were familiar with the procedures and terms used in the current ATC practice.

Apparatus

Two monitors were provided to show an ATC simulator and a CRA aid, in both ATCO and pseudo-pilot positions. A PC-based ATC simulation software, i.e. ATCSimulator®2 was used. The ATC simulator consisted of three components shown in separate windows: a 60-NM range radar display, arrivals and departures flight progress strips (Figure 1). The traffic scenario in airspace was generated with an ATC-Sector Design Kit (ATC-SDK). The flight plans including fleet times, positions, and speeds of aircraft were manipulated in order to generate conflicts.

[Insert Figure 1. about here]

A PC-based low-fidelity of a CRA aid was developed. The CRA showed the predicted conflicting pair of aircraft, the resolution maneuver advisory, and the option buttons to accept or reject the proposed resolution (see Figure 2). The list of possible abbreviations showed by the CRA aid is provided in Appendix 1. The CRA worked based on the principle for Resolution Aircraft and Maneuver Selector (RAMS) proposed by Erzberger (2006). The CRA aid applied the altitude-first resolver principle that means a vertical maneuver would be suggested first over lateral and speed maneuvers due to its expediency (Rantanen & Wickens, 2012).

In the automated conditions, the CRA provided a resolution advice two minutes prior to a conflict to alert the participants and provide a resolution advice for the predicted conflict. Neither an alert nor resolution advice was provided in the manual condition. The conflict was defined when the vertical separation between two aircraft is less than 1000 feet and the horizontal separation is less than five nautical miles. When the conflict occurred, the ATC

simulator activated a beeping sound and listed the pair of conflicting aircraft on the radar display and it would remain activated until the conflict was resolved.

[Insert Figure 2. about here]

Design

A mixed-factorial design was adopted. The first factor, automation level was a within-subjects factor with four levels: reliable, Unreliable and Short Time Allowance to Secondary conflict (U-STAS), Unreliable and Low TAS (U-LTAS), and manual (Table 2). The four automation levels were presented across participants in a sequence specified by a balanced-Latin square method. The second factor, traffic density was a between-subjects factor with two levels: medium vs. high. Sector density for medium and high traffic density was 60 and 90 aircraft, respectively. The participants were randomly assigned to these two traffic density conditions with each condition having twelve participants.

The conflict scenarios were similar across the four testing conditions. However, the aircraft call-signs, waypoints and occurrence times were changed. The traffic patterns were also rotated. This produced generally similar scenarios across the testing conditions to ensure that other factors such as conflict scenario (Thomas & Rantanen, 2006) would not influence the effects of manipulation in different testing conditions.

Table 2. Experimental Design

Condition	Reliable	U-STAS	U-LTAS	Manual
High Traffic	X	X	X	X
Low Traffic	X	X	X	X

The dependent measures were: *The task performance measures* including percentage of resolved conflicts, conflict resolution time, and aid utilization rate. Percentage of resolved conflicts was operationally defined as the absence of loss of separation (LOS) relative to the number of designated conflicts. Conflict resolution time was defined only for the automation conditions and reflected the time interval between the CRA aid onset and the pilots' maneuvering response. Aid utilization rate assessed participants' dependence on the CRA aid and reflected by the ratio of the accepted advisories relative to the total number of advisories.

NASA-TLX rating scale (total workload range from 1 to 100) (Hart & Staveland, 1988) was used to obtain a subjective *measure of mental workload*. The objective measures of mental workload were derived from ready response latency and percentage of timeouts in the Situation Present Assessment Method (SPAM) (Loft, et al., 2015). Ready response latency was defined as the interval between a ready prompt's onset and participants' response to the ready prompt. Percentage of timeouts was defined as non-responded questions.

For *the trust rating*, the Likert-type rating scale (ranging from 0 to 7) (Jian, Bisantz, & Drury, 2000) was administered after each testing condition.

Situation awareness (SA) was measured throughout the experiment using the SPAM (Durso, et al., 1999, Durso & Dattel, 2004). This method incurs fewer burdens onto ATCOs' memory and is less intrusive compared to other SA assessment tools (Bacon & Strybel, 2013) such as SAGAT (Endsley, 1988). SPAM ready response latency indicated the readiness of participants to answer SA questions and often correlates with subjective workload (Strybel, et al., 2008; Vu et al., 2012). Once participants indicated that they are ready, the accuracy and time taken to then answer the SPAM queries reflect SA (Durso & Dattel, 2004). Hence, two SA

measures were recorded (a) probe response latency defined as the interval between the display of a question and participants' answer, and (b) percentage of correct responses to SA probes.

Tasks

Each participant performed ATC tasks with the goals of maintaining separation and controlling the traffic flow. Participants used voice transmission to communicate with a pseudo-pilot. Participants had to handle all arriving and departing aircraft within their controlled area. For arriving aircraft, participants needed to: (1) accept incoming aircraft to their area, (2) clear the aircraft altitude to the landing altitude, (3) assign an appropriate approach clearance and runway, either Instrument Landing System (ILS) or Visual Approach, and (4) hand-over the aircraft to the tower controller. For departing aircraft, participants were responsible for (1) climbing the aircraft to the exit altitude, and (2) handing-over the aircraft to sector facilities.

Upon receiving voice instructions from the controllers, the pseudo-pilot inserted respective commands to the ATC simulator. The ATC simulator used in the study provides synthetic voices that met standard phraseology used in the real ATCO-pilot communication procedures when the pseudo-pilot inserted any instructed commands through keystrokes.

In the manual condition, participants were required to perform all the above tasks manually, without the CRA. In the automated conditions, the participants had available recommendations of the CRA. Six conflicts, including five pre-set primary conflicts and one secondary conflict, were imposed in the unreliable conditions. The detection of primary conflicts was 100% accurate and led to the automated resolution recommendation at 2 minutes prior to the LOS. Participants interacted with the onscreen CRA aid control using a computer mouse

connected to the system. Participants could either accept or reject the resolution advisories provided by the CRA by clicking accept or reject button, respectively.

If the ATCOs accepted the advice, the resolution advice would be automatically sent to the pseudo-pilot's screen. The pseudo-pilot would directly apply the resolution by giving the preset macro commands to the simulator. The conflict resolution would always be effective to successfully resolve the primary conflict. If the ATCOs rejected the resolution advice, the CRA would stop processing the respective aircraft' data, and would not be triggered again if the ATCOs implemented an ineffective resolution. The CRA would not have enough time to be triggered (i.e. less than 2 min) and generate a resolution for any ineffective resolution made by the ATCOs. This situation then was counted as an unresolved conflict for the conflict resolution performance.

Automation Implementation

Participants were provided with a resolution advisory for each potential conflict in the three automation conditions. In the reliable condition, all advisories provided by the CRA aid were correct (100% of aid reliability). In the unreliable conditions, the aid reliability was 80% (1 failure out of 5 advisories). The advice for the fourth conflict helped resolve the primary conflict but led to a secondary conflict with a traffic aircraft and no advice was provided to avoid the secondary conflict. We put the failure in this fourth conflict trial because it would allow participants to develop their trust in the CRA aid over the prior three trials, and for the primary conflict of the fourth trial before a failure occurred.

Our decision to impose unreliable automation by the CRA missing a secondary conflict allowed for the implementation of automation error in automation advice while experimentally

better controlling for the difference in maneuver preference between ATCOs. Furthermore, models for the development of on-ground CRA are complex and require a large number of rules to completely cover all possible encounter situations and it may fail in certain situations. (Kuchar & Yang, 2000; Rantanen & Wickens, 2012). Hence, this study evaluated the failure situation that may likely happen in future ATC environment.

In the U-STAS condition, if the ATCO accepted the CRA, then it would successfully resolve the first conflict as noted above, but this resolution would trigger a secondary conflict between one of the conflicting aircraft in the primary conflict and a traffic aircraft **100 seconds** after implementing the resolution (Figure 3). The traffic aircraft was intentionally preset in the scenario. For the secondary conflict, the CRA would not be triggered again, since this was the meaning of CRA imperfection in the study. The radar system generated an alert for the ATCO if a secondary conflict (loss of separation) occurred by a beeping sound as the evidence to the ATCO of the error.

[Insert Figure 3. about here]

In the U-LTAS condition, the secondary conflict would again be triggered between one of the aircraft in primary conflict and a traffic aircraft but now, this would be **4 minutes** after implementing the resolution maneuver (Figure 4).

[Insert Figure 4. about here]

Procedure

All participants were provided with a training session. Participants completed two 30-minutes radar training (i.e. without and with the CRA). During the practice session, participants

were told that some complex factors such as flow constraints, area boundary, and iteration process could be potential triggers for the CRA aid to err. All participants were able to successfully give clearance for all departing and arriving aircraft at the end of the training session.

During the experiment sessions, four one-hour ATC scenarios were provided corresponding to the four experimental conditions, each on a different day. Participants were instructed to perform appropriate ATC activities to deal with the simulated air traffic situations. Participants were also instructed to respond to SA ready probes and resolve SA question probes that cover all SA levels if they had more attentional resources. Each probe appeared every 6 minutes; hence there were 9 probe questions in a one-hour scenario. Seven probes appeared before the first CRA failure (i.e. the fourth conflict), 1 probe appeared during the CRA failure (in between primary conflict and secondary conflict which was within 4 minutes and 100 seconds under U-STAS and U-LTAS, respectively), and 1 probe appeared after the post failure trial. The ready and question probes would disappear after 1 minute of no response. The list of the probes is provided in Appendix 2.

Statistical analysis

A 4(automation level) x 2(traffic density) mixed-design ANOVA was performed to analyze the percentage of resolved conflicts, mental workload, and SA. The trust ratings, CR time, and dependence were analyzed using a 3x2 mixed-design ANOVA since the data were only collected in the three CRA conditions.

In addition to the omnibus analyses, we performed a series of planned orthogonal contrasts to closely examine the effects of CRA automation and TAS. First, the primary conflicts

1 to 5 were analyzed by pooling the data of all the trials to examine the effects of automation versus manual condition with the medium and high traffic densities. Second, to investigate the effects of CRA reliability, a contrast analysis (i.e. reliable vs unreliable) was performed exclusively on the fifth conflict in the automation conditions for the conflict resolution time, dependence, ready response latency (i.e. workload indicator), and probe response latency and accuracy (i.e. SA indicators). This 5th conflict was chosen because it was only here, in the two unreliable blocks, those participants would be aware that the automation was imperfect, having experienced the automation failure on the previous conflict. Hence, on this conflict, we could measure the direct effects of unreliability on dependence. Finally, a targeted analysis on the secondary conflict (occurring in trial 4) was performed to investigate the effects of TAS on the resolution performance.

RESULTS

Task performance measures

Percentage of resolved conflict

The main effect of automation level on the percentage of resolved conflicts was significant, $F(3, 66)=40.30, p<0.01, \eta^2=0.647$. All automation groups ($M=87.50\%$, $SE=4.24\%$) improved resolution performance above the manual condition where participants were not equipped with the CRA ($M=50.00\%$, $SE=5.34\%$), $F(1, 66)=103.98, p<0.01$ (Figure 5). More conflicts were successfully resolved under the medium traffic density ($M=87.08\%$, $SE=4.06\%$) than high traffic density ($M=69.17\%$, $SE=4.97\%$) conditions, $F(1, 22)=17.55, p<0.01, \eta^2=0.444$. A significant interaction effect between automation level and traffic density was observed, $F(3,$

66)=3.03, $p=0.04$, $\eta^2=0.121$, confirmed the amplification of CRA aid benefits under high traffic. Thus both H1 and H2 were strongly confirmed.

For the fifth conflict, there was no performance difference between reliable ($M=95.93\%$, $SE=4.17\%$) and unreliable ($M=89.58\%$, $SE=4.46\%$) conditions, $F(1,44)=0.85$, $p=0.36$, indicating the benefit of even CRA automation that had proven to be unreliable in the previous conflict.

For the targeted analysis of the secondary conflict, no significant difference between STAS ($M=54.17\%$, $SE=10.39$) and LTAS ($M=70.83\%$, $SE=9.48\%$) was found, $F(1,22)=1.09$, $p=0.31$, although it is noted that both of these values are considerably lower than the average over all CRA-supported trials, reflecting the costs of an automation failure.

[Insert Figure 5. about here]

Conflict resolution time

Conflicts were resolved significantly faster under the medium ($M=21.33s$, $SE=3.12$) than under high traffic density ($M=38.33s$, $SE=4.95$) $F(1,22)=15.79$, $p<0.01$, $\eta^2=0.418$. There were no other significant results on resolution time.

Aid utilization rate

For the fifth conflict, no difference in dependence between reliable ($M=79.17\%$, $SE=8.47\%$) and unreliable ($M=75.55\%$, $SE=7.37\%$) conditions was found, $F(1,44)=0.17$, $p=0.68$. Higher dependence was observed under U-LTAS ($M=87.5\%$, $SE=6.89\%$) than U-STAS ($M=62.5\%$, $SE=10.09\%$), $F(1,44)=4.57$, $p=0.04$. There were no other significant results.

Mental workload vs Subjective ratings of mental workload

Ready response latency in the SPAM indexed objective workload. There was a significant difference for ready response latency to suggest that automation ($M=9.25s$, $SE=1.68s$) did reduce workload than manually performing the tasks ($M=13.58s$, $SE=2.44s$) by shortening this latency, $F(1, 66)=7.50$, $p=0.01$. For the subjective workload measure, there were no significant results.

For the fifth conflict, higher workload as indicated by longer ready latency was found under U-LTAS ($M=11.48s$, $SE=2.63s$) than U-STAS ($M=5.95s$, $SE=0.84s$) condition, $F(1, 44)=5.25$, $p=0.03$.

Situation awareness

No differences between automated and manual conditions were found on probe accuracy ($M=70.21\%$, $SE=4.57\%$ vs $M=66.80\%$, $SE=5.07\%$), $F(1, 66)=1.01$, $p=0.32$ as well as on probe latency ($M=14.42s$, $SE=0.73s$ vs $M=15.08s$, $SE=1.03s$), $F(1, 66)=0.80$, $p=0.37$. There were no significant effects of traffic density on probe accuracy ($M_{High}=67.39\%$, $SE=7.85\%$ vs $M_{Low}=71.69\%$, $SE=5.25\%$), $F(1, 22)=0.42$, $p=0.52$ as well as on probe latency ($M_{High}=14.23s$, $SE=0.97s$ vs $M_{Low}=14.94s$, $SE=1.29s$), $F(1, 22)=0.29$, $p=0.59$.

For the fifth conflict, there were no differences between reliable and unreliable conditions in probe accuracy ($M=88.89\%$, $SE=5.68\%$ vs $M=87.09\%$, $SE=5.86\%$), $F(1, 44)=0.05$, $p=0.82$ as well as in probe response latency ($M=15.43s$, $SE=0.81s$ vs $M=18.52s$, $SE=1.84$), $F(1, 44)=2.17$, $p=0.15$.

Subjective ratings of trust

Trust was higher after the blocks in which participants were provided with reliable (M=5.54, SE=0.21) than unreliable CRA (M=5.23, SE=0.23), $F(1, 44)=3.99, p=0.05$. There were no other significant results.

DISCUSSION

The current experiment addressed the extent to which an imperfect ATC CRA aid could assist controller performance, and provided results relative to the more general theory of HAI. The hypotheses and the results are summarized in Table 3.

Our first hypothesis was confirmed regarding performance and objective workload, replicating the prior findings of Trapsilawati et al. (2015) but here at the much higher traffic load of twice and three times the volume. Participants used (depended) on the automation, and their maneuver resolution performance with the aid was substantially above unaided performance. Furthermore, workload was reduced by automation, since ATCOs' ability to effectively and safely maintain aircraft separation depends upon the maintenance of their mental picture (Nunes & Mogford, 2003; Alexander, Wickens & Merwin, 2005); the CRA could reduce the amount of mental computation thus reduce their workload. We also observed increased CRA benefit in resolving conflicts under higher (47.8%) than under lower traffic load (27.2%), supporting Hypothesis 2.

Regarding Hypothesis 3, the analysis on trial 5, at which time, from the participant's perspective, automation was now seen to be 80% reliable in the two unreliable conditions, no decrement in performance was observed relative to the 100% reliable condition, hence providing additional data in support of the conclusions of Wickens & Dixon (2007) that reliability as low

as 80% does not impose substantial costs, and thereby generalizing their conclusions from automated detection tasks, to those of automated decision aids.

Table 3. The summary of the hypotheses result

Hyp. No	Hypotheses	Results
H1	Imposing the CRA automation would increase both performance and reduce workload (when compared to fully manual conflict resolution), given that controllers use (depend upon) the automation.	Confirmed
H2	The benefits in (1) would be amplified under high traffic load, when cognitive resources were scarcer.	Confirmed
H3	The benefits of the CRA automation would be manifest even when it was imperfectly reliable, given that its reliability was greater than the threshold of automation assistance between 0.70 & 0.75 (Wickens & Dixon, 2007).	Confirmed
H4	CRA automation, whether perfect or imperfect would decrease situation awareness (SA) of the traffic situation.	Not confirmed
H5	The effects in (4) would be amplified under high workload (high traffic load).	Not confirmed
H6	After automation had failed, the trust in the CRA would decline but the dependence would not decline, due to the high traffic load.	Confirmed
H7	Shorter TAS should decrease situation awareness, increase workload and increase dependence on automation.	Not confirmed

Importantly automation did not lower situation awareness, in contrast to the prototypical model of automation dependency (Higham, et al., 2013; Onnasch, Ruff, et al., 2014; Onnasch, Wickens, et al., 2014). Our null finding here failed to confirm H4, but replicated the findings of Trapsilawati, et al. (2015) regarding CRA aid benefits to overall situation awareness and again is consistent with the assumption that our CRA aid-supported participants continued high engagement in the traffic situation. Thus, any possible costs of imposing automation to situation awareness appeared to have been offset by its workload reduction benefits, which availed more resources for monitoring the raw traffic data. The absence of an interaction of automation with traffic load on SA also failed to confirm H5.

The effects of CRA on trust and dependence were in the expected direction. Trust was degraded in the unreliable automation blocks. However, dependence on the CRA was not degraded after it had failed, supporting Hypothesis 6. This occurred due to high workload so that participants opted to rely on the CRA to preserve cognitive resources for other ATC tasks although they did not trust it as much.

Regarding H7, the effects of TAS in this experiment were muted, just as they were in Trapsilawati, et al. (2015). TAS did not affect SA. This might be because even the short TAS occurred within their cognitive capacity and reflected conflict alert in the current ATC operations, since the CRA here represented the integration of conflict detection and resolution system, participants might be familiar with this condition thus not compromises their SA. In contrast to H7, shorter TAS did not increase dependence (utilization), but in fact decreased it. Correspondingly, shorter TAS did not increase workload, but also decreased that variable, when measured by ready response time. Thus these two dependent variables were linked in an interpretable way: when workload was inferred to be higher, dependence increased; but the cause

of the increase in workload was, unexpectedly, a longer, not a shorter TAS. We may interpret this somewhat unexpected result by assuming that longer TAS allowed participants more time and resources to attend to the raw radar data, and use their unaided cognitive resources to help choose the appropriate maneuver. Greater reliance on their own cognition is a very plausible source of higher workload.

Limitations and implications

There exist some limitations in the present study. First, the ATC simulator used in the present study may not fully represent the real ATC environment. Real world stressors such as weather changes were not taken into account. Second, most of the participants were not active ATCOs. But, our participants all had ATC training and knowledge and were familiar with ATC system through prior training. Further, both ATCOs and students have been found in previous studies to have similar preference in resolution maneuvering instruction patterns (Rantanen & Nunes, 2005). Since the CRA has not been applied in the current ATC workplace, this research was conducted to gather preliminary information in the CRA context. Using students as participants well fitted with the explanatory nature of the present study (Goritzlehner et al., 2014). Third, the conflict alerting function integrated into the CRA was always reliable while previous studies indicated different performance consequences due to misses and false alarms of conflict detection algorithms (Rovira & Parasuraman, 2010; Dixon, Wickens, & McCarley, 2007; Wickens & Colcombe, 2007). Thus, further research incorporating this issue is required. Lastly, the analysis of single “automation failure” trials, must of necessity be of considerably lower statistical power than the overall performance measures because, by definition, unexpected

failures occur rarely (Wickens, 2009). Hence, we must interpret more cautiously those null effects that are observed on the single conflict trial (4 or 5).

The present study has positive implications for CRA aid use. First, the results implied that system designers must put high efforts in providing a reliable automation's decision in the first conflict resolution iteration to turn the "totally unexpected" effects to the "surprising" effects. For the air traffic management industry, this implication shifts the complexity of CRA development to a feasible level. As of today, research is still ongoing for the development of CRA; however, considering every single constraint in the airspace to provide a fully secondary conflict-free resolution is nearly impossible (Kuchar & Yang, 2000). Yet, ATCOs are waiting for the real support in facing the imminent traffic growth. Next, we found that CRA aid could improve ATCOs' performance and reduce workload across traffic levels, verifying its applicability across traffic densities. Moreover, participants worked well with the CRA aid and preferred to share the responsibility of separation assurance (Cabrall et al., 2014) with the CRA aid. Thus, setting automation at the moderate level (Bekier, Molesworth, & Williamson, 2012; Li, et. al, 2014; Parasuraman, et al., 2000) that supports rather than replaces ATCOs (Prevot, et al. 2012) was recommended in developing the CRA aid. This allows ATCOs to complement the CRA aid safely with their manual intervention, should any automation failure occurs.

KEY POINTS

- The human-automation interaction issues of imperfect automation and other underlying factors in future air traffic control were examined in the context of conflict detection and resolution system.

- 454 • CRA aid benefits ATCOs, notwithstanding its possible imperfection, in different traffic
455 densities (27.22% and 47.78% improvement in lower and higher traffic load, respectively, as
456 compared to manually performing the task).
- 457 • The benefits of imperfect CRA was associated with the examination of incorrect resolution
458 with traffic aircraft which turned the “totally unexpected” event into the “surprising” event.
- 459 • TAS did not affect SA. Longer TAS led to higher workload, and thus higher dependence on
460 the CRA.
- 461

REFERENCES

- Airbus. (2013). *Future journey 2013-2032*. Airbus global market forecast report. Available from http://www.airbus.com/company/market/forecast/?eID=dam_frontend_push&docID=337 52.
- Alexander, A. L., Wickens, C. D., & Merwin, D. H. (2005). Perspective and coplanar cockpit displays of traffic information: Implications for maneuver choice, flight safety, and mental workload. *The International Journal of Aviation Psychology*, 15, 1-21.
- Bacon, L. P., & Strybel, T. Z. (2013). Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. *Safety Science*, 56, 89-95.
- Bekier, M., Molesworth, B. R., & Williamson, A. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety Science*, 50, 259-265.
- Cabrall, C., Gomez, A., Homola, J., Hunt, S., Martin, L., Mercer, J., & Prevot, T. (2014). Transitioning Resolution Responsibility between the Controller and Automation Team in Simulated NextGen Separation Assurance. In S. Ozeki (Ed.), *Air Traffic Management and Systems* (pp: 147-172). Tokyo, Japan: Springer.
- Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, 49, 564-572.
- Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137-154). Hampshire, UK: Ashgate.

484 Durso, F. T., Hackworth, C., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998).
 485 Situation awareness as a predictor of performance in en route air traffic controllers. *Air*
 486 *Traffic Control Quarterly*, 5, 1–20.

487 Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT).
 488 In *Proceedings of the IEEE National Conference on Aerospace and Electronics* (pp. 789-
 489 795). Dayton, OH: IEEE.

490 Erzberger, H. (2006). Automated conflict resolution for air traffic control. In *Proceedings of the*
 491 *25th International Congress of the Aeronautical Sciences* (pp. 1-28). Hamburg,
 492 Germany: International Council of the Aeronautical Sciences.

493 Goritzlehner, R., Borst, C., Ellerbroek, J., Westin, C., van Paassen, M., & Mulder, M. (2014).
 494 Effects of transparency on the acceptance of automated resolution advisories. In
 495 *Proceedings of the 2014 IEEE International Conference on Systems, Man and*
 496 *Cybernetics* (pp. 2965-2970). San Diego, CA: Institute of Electrical and Electronics
 497 Engineers.

498 Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (Task Load Index): Results
 499 of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human*
 500 *Mental Workload* (pp. 139-183). Amsterdam, Netherlands: Elsevier.

501 Higham, T.-L. M., Vu, K.-P. L., Miles, J., Strybel, T. Z., & Battiste, V. (2013). Training air
 502 traffic controller trust in automation within a NextGen environment. In S. Yamamoto
 503 (Ed.), *HIMI/HCI 2013* (Vol. 8017, pp. 76-84). Heidelberg: Springer.

504 Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined
 505 scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4,
 506 53-71.

507 Kuchar, J. K., & Yang, L. C. (2000). A review of conflict detection and resolution modeling
508 methods. *IEEE Transactions on Intelligent Transportation Systems*, 1, 179-189.

509 Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and Levels of Automation in
510 Support of Space Teleoperations. *Human Factors*, 56, 1050-1061.

511 Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015) Situation
512 awareness measures for simulated submarine track management. *Human Factors*, 57,
513 298-310.

514 Manzey, D., Gérard, N., & Wiczorek, R. (2014). Decision-making and response strategies in
515 interaction with alarms: the impact of alarm reliability, availability of alarm validity
516 information and workload. *Ergonomics*, 57, 1833-1855.

517 Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of
518 decision aid reliability on controller performance and mental workload. *Human Factors*,
519 47, 35-49.

520 Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure:
521 Vigilance and task complexity effects. *Human Factors*, 38, 311-322.

522 Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids:
523 Made for each other. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human*
524 *Performance: Theory and applications* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum
525 Associates.

526 Nunes, A., & Mogford, R. H. (2003). Identifying controller strategies that support the 'Picture'.
527 In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp.
528 71-75). Denver, CO: Human Factors and Ergonomics Society.

529 Onnasch, L., Ruff, S., & Manzey, D. (2014). Operators' adaptation to imperfect automation—
 530 impact of miss-prone alarm Systems on attention allocation and performance.
 531 *International Journal of Human-Computer Studies*, 72, 772-782.

532 Onnasch, L., Wickens, C., Li, H., & Manzey, D. (2014). Human performance consequences of
 533 stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56, 476–
 534 488.

535 Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation:
 536 An attentional integration. *Human Factors*, 52, 381-410.

537 Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-
 538 induced 'complacency'. *The International Journal of Aviation Psychology*, 3, 1-23.

539 Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Level of
 540 Human Interaction with Automation. *IEEE Transactions on Systems, Man, and*
 541 *Cybernetics-Part A: Systems and Humans*, 30, 286-289.

542 Prevot, T., Homola, J., & Mercer, J. (2008). Human-in-the-loop evaluation of ground-based
 543 automated separation assurance for NEXTGEN. In *Proceedings of the 26th International*
 544 *Congress of the Aeronautical Sciences*, AIAA-ATIO-2008-8885.

545 Prevot, T., Homola, J. R., Martin, L. H., Mercer, J. S., & Cabrall, C. D. (2012). Toward
 546 Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in
 547 Human/Systems Interaction. *International Journal of Human-Computer Interaction*, 28,
 548 77-98.

549 Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *The*
 550 *International Journal of Aviation Psychology*, 15, 339-362.

551 Rantanen, E. M., & Wickens, C. D. (2012). Conflict Resolution Maneuvers in Air Traffic
552 Control: Investigation of Operational Data. *The International Journal of Aviation*
553 *Psychology*, 22, 266-281.

554 Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision
555 making in a simulated command and control task. *Human Factors*, 49, 76-87.

556 Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of
557 imperfect automation on controller attention and performance. *Human Factors*, 52, 411-
558 425.

559 Strybel, T. Z., Vu, K. L., Kraft, J., & Minakata, K. (2008). Assessing the situation awareness of
560 pilots engaged in self spacing. In *Proceedings of the Human Factors and Ergonomics*
561 *Society 52nd Annual Meeting* (pp. 11–15). Santa Monica, CA: Human Factors and
562 Ergonomics Society.

563 Thomas, L. C., & Rantanen, E. M. (2006). Human factors issues in implementation of advanced
564 aviation technologies: A case of false alerts and cockpit displays of traffic information.
565 *Theoretical Issues in Ergonomics Science*, 7, 501-523.

566 Trapsilawati, F., Qu, X., Wickens, C. D., & Chen, C.-H. (2015). Human factors assessment of
567 conflict resolution aid reliability and time pressure in future air traffic control.
568 *Ergonomics*, 58, 897-908.

569 Vossen, T. W. M, Hoffman, R, & Mukherjee, A. (2012). Air traffic flow management. In C.
570 Barnhart & B. Smith (Eds.), *Quantitative problem solving methods in the airline industry*
571 (pp. 385-453). US: Springer.

572 Vu, K. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A. V., Brandt, S., Ligda, S., & Johnson,
573 W. (2012). Pilot performance in trajectory-based operations under concepts of operation

574 that vary separation responsibility across pilots, air traffic controllers, and
575 automation. *International Journal of Human-Computer Interaction*, 28, 107–118.

576 Wickens, C. D. (2009). The psychology of aviation surprise: an 8 year update regarding the
577 noticing of black swans. In J. Flach & P. Tsang (eds). In *Proceedings of the 2009*
578 *Symposium on Aviation Psychology* (Keynote address). Dayton, OH: Wright State
579 University.

580 Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and
581 automation bias in the use of imperfect automation. *Human Factors*, 57, 728-739.

582 Wickens, C. D., & Colcombe, A. (2007). Dual-task performance consequences of imperfect
583 alerting associated with a cockpit display of traffic information. *Human Factors*, 49, 839-
584 850.

585 Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A
586 synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8, 201-212.

587 Wickens, C. D., Hollands, J. G., Banburry, S., & Parasuraman, R. (2013). *Engineering*
588 *Psychology and Human Performance*. New Jersey, US: Pearson.

589 Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying
590 black swans in NextGen: Predicting human performance in off-nominal conditions.
591 *Human Factors*, 51, 638-651.

592 Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False Alerts
593 in Air Traffic Control Conflict Alerting System: Is There a “Cry Wolf” Effect? *Human*
594 *Factors*, 51, 446-462.

595 Westin, C. A., Borst, C., & Hilburn, B. (2013). Mismatches between automation and human
596 strategies: an investigation into future air traffic management decision aiding. *In*

597 *Proceedings of the International Symposium on Aviation Psychology* (pp. 1-6). Ohio,
598 USA.

599 Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down:
600 The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display
601 signaling. *Human Factors*, 45, 390-407.

602

603

604 **Appendix 1: List of maneuvering abbreviations and the units**

Notation	Abbreviation	Unit
C	Climb	Flight level
D	Descend	Flight level
FH	Fly heading	Degree
IS	Increase speed	Knots
RS	Reduce speed	Knots

605

606 **Appendix 2: List of probe questions**

Level 1	
1	What is the aircraft A's speed?
2	What is the direction of departure for aircraft B?
3	What is the altitude clearance for aircraft C?
Level 2	
1	How many aircraft flying southbound?
2	Which aircraft has lower altitude?
3	Is the difference in heading between aircraft D and aircraft E more than 90°?
Level 3	
1	Which aircraft must be handed-off to another sector within the next 2 minutes?
2	Which pairs of aircraft will lose separation if they stay on their current courses?
3	Which aircraft will need a new clearance to achieve landing requirements?

607

608

609 **Figure Captions**

610 **Figure 1.** The display of the ATC simulator.

611 **Figure 2.** Conflict resolution aid. In this example, the pilot of aircraft N755GH was advised to
612 climb to fourteen thousand feet (flight level 140) while the aircraft N74932 maintained
613 its current course.

614 **Figure 3.** The U-STAS condition.

615 **Figure 4.** The U-LTAS condition.

616 **Figure 5.** Percentage of resolved conflicts across traffic densities (error bars indicate 1 SE).

617

618 **BIOGRAPHIES**

619 Fitri Trapsilawati is a PhD candidate at the School of Mechanical and Aerospace Engineering,
620 Nanyang Technological University, Singapore. She earned her B.Eng in Industrial Engineering
621 from the Universitas Gadjah Mada, Indonesia in 2010.

622

623 Christopher D. Wickens is a professor emeritus of aviation and psychology at the University of
624 Illinois and is currently a senior scientist at Alion Science and Technology, Boulder, Colorado
625 and Professor of Psychology at Colorado State University.

626

627 Xingda Qu is a professor at Shenzhen University, Shenzhen, China. He is also serving as the
628 director of the Institute of Human Factors and Ergonomics at Shenzhen University. He earned
629 his PhD in Industrial Engineering from Virginia Tech, United States in 2008.

630

631 Chun-Hsien Chen is an associate professor, and the director of the design stream, at the School
632 of Mechanical & Aerospace Engineering, Nanyang Technological University, Singapore. He
633 earned his PhD in Industrial Engineering from the University of Missouri-Columbia, United
634 States.